Iniziato	Thursday, 13 January 2022, 15:13
Stato	Completato
Terminato	Thursday, 13 January 2022, 15:30
Tempo impiegato	17 min. 13 secondi
Punteggio	15,00/15,00
Valutazione	**30,00** su un massimo di 30,00 (**100**%)

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main reason for the *MinMax scaling* (also known as "*rescaling*") of attributes?

Scegli un'alternativa:

a. Map all the numeric attributes to the same range, in order to prevent the attributes with higher range from having prevalent influence ✔

b. Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence

c. Remove abnormal values

d. Change the distribution of the numeric attributes, in order to obtain gaussian distributions

Your answer is correct.

La risposta corretta è: Map all the numeric attributes to the same range, in order to prevent the attributes with higher range from having prevalent influence

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the two binary vectors below, which is their similarity according to the Jaccard Coefficient?

a b c d e f g h i j

1 0 0 0 1 0 1 1 0 1

1 0 1 1 1 0 1 0 1 0

Scegli un'alternativa:

- a. 0.1
- b. 0.5
- c. 0.2
- d. 0.375 ✔ 3/8 is the fraction of matching 1's, divided by (the number of matching 1 plus the number of non-matching)

Risposta corretta.

It is the number of matching 1 divided by the number of matching 1 + the number of non-matching

La risposta corretta è: 0.375

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the *single linkage?*

Scegli un'alternativa:

- a. A method to compute the separation of the objects inside a cluster
- b. A method to compute the distance between two objects, it can be used in hierarchical clustering
- c. A method to compute the distance between two sets of items, it can be used in hierarchical clustering ✔
- d. A method to compute the distance between two classes, it can be used in decision trees

Your answer is correct.

La risposta corretta è: A method to compute the distance between two sets of items, it can be used in hierarchical clustering

Domanda **4**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the *precision* of a binary classifier?

Scegli un'alternativa:

a. TN / (TN + FP)

b. TP / (TP + FP) ✔ This is also called *positive predictive value*, which is the number of detected true positives divided by the total number of elements predicted as positive

c. (TP + TN) / (TP + FP + TN + FN)

d. TP / (TP + FN)

Risposta corretta.

La risposta corretta è: TP / (TP + FP)

Domanda **5**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Why do we *prune* a decision tree?

Scegli un'alternativa:

a. To eliminate rows of the dataset which could be influenced by random effects

b. To eliminate parts of the tree where the decisions could be influenced by random effects ✔

c. To eliminate parts of the tree where the decision could generate *underfitting*

d. To eliminate attributes which could be influenced by random effects

Your answer is correct.

La risposta corretta è: To eliminate parts of the tree where the decisions could be influenced by random effects

Domanda **6**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In a decision tree, an attribute which is used only in nodes near the leaves...

Scegli un'alternativa:

- a. ...gives little insight with respect to the target ✔
- b. ...is irrelevant with respect to the target
- c. ...has a high correlation with respect to the target
- d. ...guarantees high increment of purity

Risposta corretta.

La risposta corretta è: ...gives little insight with respect to the target

Domanda **7**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What are the hyperparameters of a Neural Network? (Possibly non exhaustive)

- a. Input layers structure, Learning rate, Activation function, Number of epochs
- b. Hidden layers structure, Learning rate, Activation function, Number of epochs ✔
- c. Network structure, Learning rate, Backpropagation algorithm, Number of epochs
- d. Hidden layers structure, Output layer structure, Activation function, Number of epochs

Your answer is correct.

La risposta corretta è:
Hidden layers structure, Learning rate, Activation function, Number of epochs

Domanda **8**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following is a strength of the clustering algorithm DBSCAN?

Scegli una o più alternative:

- a. Very fast computation
- b. Ability to find cluster with concavities ✔
- c. Ability to separate outliers from regular data ✔
- d. Requires to set the number of clusters as a parameter

Your answer is correct.

Le risposte corrette sono: Ability to find cluster with concavities, Ability to separate outliers from regular data

Domanda **9**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What does K-means try to minimise?

Scegli un'alternativa:

- a. The *separation*, that is the sum of the squared distances of each point with respect to its centroid
- b. The *distortion*, that is the sum of the squared distances of each point with respect to the points of the other clusters
- c. The *distortion*, that is the sum of the squared distances of each point with respect to its centroid ✔
- d. The *separation*, that is the sum of the squared distances of each cluster centroid with respect tho the global centroid of the dataset

Risposta corretta.

La risposta corretta è: The *distortion*, that is the sum of the squared distances of each point with respect to its centroid

Domanda **10**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the statements below is true? (One or more)

Scegli una o più alternative:

- [x] a. Increasing the radius of the neighbourhood can decrease the number of noise points ✔
- [ ] b. DBSCAN always stops to a configuration which gives the optimal number of clusters
- [x] c. DBSCAN can give good performance when clusters have concavities ✔
- [x] d. Sometimes DBSCAN stops to a configuration which does not include any cluster ✔

Your answer is correct.

Le risposte corrette sono: Sometimes DBSCAN stops to a configuration which does not include any cluster, DBSCAN can give good performance when clusters have concavities, Increasing the radius of the neighbourhood can decrease the number of noise points

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Match the rule evaluation formulas with their names

$$\frac{conf(A \Rightarrow C)}{sup(C)}$$ Lift ✔

$$\frac{sup(A \Rightarrow C)}{sup(A)}$$ Confidence ✔

$$sup(A \cup C) - sup(A)sup(C)$$ Leverage ✔

$$\frac{1 - sup(C)}{1 - conf(A \Rightarrow C)}$$ Conviction ✔

Your answer is correct.

La risposta corretta è: $\frac{conf(A \Rightarrow C)}{sup(C)}$ → Lift, $\frac{sup(A \Rightarrow C)}{sup(A)}$ → Confidence, $sup(A \cup C) - sup(A)sup(C)$ → Leverage, $\frac{1 - sup(C)}{1 - conf(A \Rightarrow C)}$ → Conviction

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Consider the transactional dataset below

ID	Items
1	A,B,C
2	A,B,D
3	B,D,E
4	C,D
5	A,C,D,E

Which is the *support* of the rule A,C ⇒ B?

Scegli un'alternativa:

- a. 40%
- b. 100%
- c. 20% ✔ 1 / 5
- d. 50%

Risposta corretta.

La risposta corretta è: 20%

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following ***is not*** an objective of feature selection

Scegli un'alternativa:

- a. Avoid the *curse of dimensionality*
- b. Reduce time and memory complexity of the learning algorithms
- c. Select the features with higher range, which have more influence on the computations ✔
- d. Reduce the effect of noise

Risposta corretta.

La risposta corretta è: Select the features with higher range, which have more influence on the computations

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the difference between classification and regression?

- a. Classification is a supervised activity, while regression in unsupervised
- b. Classification has a categorical target, while regression has a numeric target ✔
- c. Classification can have a numeric or categorical target, while regression has always a categorical target
- d. Classification can make errors, while regression is always exact

Your answer is correct.

La risposta corretta è:
Classification has a categorical target, while regression has a numeric target

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What does K-means try to minimise?

Scegli un'alternativa:

- a. The *distortion*, that is the sum of the squared distances of each point with respect to its centroid ✔
- b. The *distortion*, that is the sum of the squared distances of each point with respect to the points of the other clusters
- c. The *separation*, that is the sum of the squared distances of each point with respect to its centroid
- d. The *separation*, that is the sum of the squared distances of each cluster centroid with respect tho the global centroid of the dataset

Risposta corretta.

La risposta corretta è: The *distortion*, that is the sum of the squared distances of each point with respect to its centroid

Vai a...

Machine Learning - Python Lab ►